

Cairo University

Faculty of Engineering

Department of Computer Engineering

**LA**خ**SLY**



A Graduation Project Report Submitted

to

Faculty of Engineering, Cairo University

in Partial Fulfillment of the requirements of the degree

of

Bachelor of Science in Computer Engineering.

**Presented by**

Ahmed Salama Hamed Ahmed Maher

Moamen hassan attia Mohamed Talaat Mohamed

**Supervised by**

Dr. MagdaFayek

2020

All rights reserved. This report may not be reproduced in whole or in part, by photocopying or other means, without the permission of the authors/department

**General Guidelines**

* The color of the cover page should be green
* Throughout your text, use the font type, font size, and spacing, as in this template. In general, Arial font should be used. Chapter headings should be of size 24. Sections should be of size 18, and regular text should be of size 12. Your text should be justified on both left and right sides.
* The reference list should be written using a font size of 10. Ensure that the references are written correctly and all fields are included. References should be ordered according to their appearance in the text “[1], [2], [3] … etc”
* The table of content is a tentative one. You could add more sections as required. However, the mentioned sections should be included in your report
* For the appendices, add any appendix you see necessary. Remove any appendix that is not applicable to your project. However, the feasibility study and user guide should be included
* Ensure that the report is clear and self-contained, such that any future interested reader could completely understand your project “to the extent of building another one similar to yours”
* Use figures as much as possible to clarify and enrich your discussion. You have to draw all figures yourself. Ensure that the figures are clear and their size is suitable.
* Any figure caption should be inserted below the figure. Figures within any chapter should be numbered starting from 1. For example, the first figure of chapter 2 should be “Figure 2.1”. Similarly, the fourth figure of chapter 3 should be “Figure 3.4”
* Any table caption should be inserted above the table. Tables within any chapter should be numbered starting from 1. For example, the first table of chapter 4 should be “Table 4.1”. Similarly, the seventh table of chapter 5 should be “Table 5.7”
* Copy and paste from any other source is not allowed by any shape. Even for the background knowledge, you have to use your own wording.
* The complete report should be submitted 48 hours before the final project demonstration day. Ensure that you would meet this deadline to avoid any late penalty

**Abstract**

Soccer is one of the most popular team sports all over the world. Most sports games are naturally organized into successive and alternating plays of offence and defence, Resulting in events such as goals or dangerous attacks. If a sport videos can be segmented according to these meaningful events, it then can be used in various applications to enhance their values and enrich the user's viewing experiences and also save the user’s time. According to this, soccer video summarization and analysis has recently attracted much research and a wide spectrum of possible applications have been considered.

Soccer video summarization and analysis is concerned with the extraction of valuable events by efficient and effective processing of combination of visual, audio and text information. However, one of the major limitations of current soccer analysis is the gap between the low level features such as (color, texture, shape and motion) and high level representation such as (shot types, shot length, and shot replays).

This thesis presents an automatic soccer video summarization system using a variety of techniques in the fields of image processing, machine learning. The proposed system is composed of five phases. Namely; in the pre-processing phase, the system reads the video stream into patches of frames. Then, in the shot Boundary phase is applied on every patch of frames to segment the video into seprate shots i.e. finding the transition from a camera into another. Then each shot is processed seperatly in the following three phases, the shot classification phase in concerned with classifiying each shot for example a wide shot is when the whole pitch is visible, A close shot is when the players faces are visible etc. then the event detection phase in which each shot is analyside to determine if a goal happens in it or not or if it the audio level in this shot is high with respect to the whole video or if the goal mouth appears in the shot . And lastly, the summarization phase in which the output of all previous phases is analysized to determine which shot is important and should include in the final video and which shots to discard

**الملخص**

تعد كرة القدم واحدة من أكثر الرياضات الجماعية شعبية حول العالم. معظم الالعاب الرياضية تكون منظمة بطبيعتها بشكل متعاقب ومتناوب للهجوم والدفاع ، مما يؤدي إلى أحداث ممتعة و مهمة مثل الأهداف أو الهجمات الخطيرة. إذا كان من الممكن تقسيم مقاطع الفيديو الرياضية وفقًا لهذه الأحداث ذات المغزى ، فيمكن استخدامها في تطبيقات مختلفة لتحسين قيمها وإثراء تجارب المشاهدة للمستخدم وتوفير وقت المستخدم أيضًا. وفقًا لذلك ، اجتذب تلخيص و تحليل مقاطع فيديو كرة القدم مؤخرًا الكثير من الأبحاث وقد تم النظر في مجموعة واسعة من التطبيقات الممكنة.

يختص تلخيص وتحليل فيديو كرة القدم باستخراج الأحداث القيمة من خلال المعالجة الفعالة والدقيقة لمجموعة من المعلومات المرئية والصوتية والنصية. ومع ذلك ، فإن أحد القيود الرئيسية لتحليل كرة القدم الحالية هو الفجوة بين الميزات منخفضة المستوى مثل (اللون ، والملمس ، والشكل والحركة) والتمثيل عالي المستوى مثل (أنواع اللقطات ، وطول اللقطة ، وإعادة اللقطات).

تقدم هذه الرسالة نظام تلخيص اوتوماتيكي لفيديو كرة القدم باستخدام مجموعة متنوعة من التقنيات في مجالات معالجة الصور ، وتعلم الاله. يتكون النظام المقترح من خمس مراحل. اولا، مرحلة ما قبل المعالجة ، يقرأ النظام الفيديو في شكل صور منفردة. ثانيا، يتم تطبيق مرحلة تحديد مدة اللقطة على كل مجموعة من الصور لتقسيم الفيديو إلى لقطات منفصلة ، أي العثور على الانتقال من كاميرا إلى أخرى. ثم يتم معالجة كل لقطة بشكل منفصل في المراحل الثلاث التالية ، مرحلة تصنيف اللقطة، معنية بتصنيف كل لقطة على سبيل المثال اللقطة الواسعة عندما تكون رقعة ملعب كرة القدم مرئية بالكامل ، اللقطة القريبة عندما تكون وجوه اللاعبين مرئية وما إلى ذلك. ثم مرحلة اكتشاف احداث المباراة و التي يتم تحليل كل لقطة فيها لتحديد ما إذا كان هناك هدف يحدث أم لا أو إذا كان مستوى الصوت في هذه اللقطة مرتفعًا مقارنة بالفيديو بأكمله أو إذا ظهر المرمي في اللقطة. وأخيرًا ، مرحلة التلخيص التي يتم فيها تحليل مخرجات جميع المراحل السابقة لتحديد أي لقطة مهمة ويجب تضمينها في الفيديو النهائي وأي اللقطات يجب تجاهله، و يتم ايضا في هذه المرحلة تصنيف الحدث المهم الذي تضمنه مجموعة من اللقطات لمعرفة ما اذا كان الحدث هو احراز هدف، ام هجمة خطيرة، ام حدث اخر مهم كالتسلل، والعركلة، وما الي ذلك.

**ACKNOWLEDGMENT**

We would like to express our sincere gratitude to Dr. Magda fayek for letting us

work on the exciting topic of sports video processing and giving us so much freedom to

explore and investigate new areas of video processing, providing invaluable personal and

professional guidance, being accessible all the time, and also convey our love for our parents, family and friends

**Table of Contents**

Abstract ...........................................................................................................…........................ iii

الملخص ........................................................................................................................................... iv

Acknowledgment ......................................................................................................................... v

Table of Contents .....................................................................................................................… vi

List of Figures ............................................................................................................................. ix

List of Tables .............................................. .............................................................................. x

List of Abbreviations .............................................. ..................................................................... xi

List of Symbols ............................................................................................................................ xii

Contacts ..................................................................................................................................... xiii

**Chapter 1: Introduction**

1.1. Motivation and Justification…..……………………………………………..……………………1

1.2. Project Objectives and Problem Definition………………………………….………………….1

1.3. Project Outcomes ............................................................................................................... 1

1.4. Document Organization ...................................................................................................... 1

**Chapter 2: Visibility Study**

2.1. Target Customers .............................................................................................................. 2

2.2. Market Survey ……………………......................................................................................... 2

2.2.1. Example of Subsection 1 .............................................................................................. 2

2.2.2. Example of Subsection 2 ............................................................................................. 2

2.2.2.1. Example of subsubsection .................................................................................. 2

2.3 Business Case and Financial Analysis ….………………………………………………………. 2

**Chapter 3: Literature Survey**

3.1. Background on Topic1 …..................................................................................................... 4

3.2. Background on Topic 2 .........................................................................................................4

3.3. Comparative Study of Previous Work....................................................................................4

3.4. Implemented Approach .............................................................................. 4

**Chapter 4: System Design and Architecture**

4.1. Overview and Assumptions ................................................................................................ 5

4.2. System Architecture ………………………………...………………………………………....… 5

4.2.1. Block Diagram .............................................................................................................. 5

4.3. Module 1 …………………………………………………………………………………..……… 6

4.3.1. Functional Description .................................................................................................. 6

4.3.2. Modular Decomposition ................................................................................................ 6

4.3.3. Design Constraints ....................................................................................................... 6

4.3.4. Other Description of Module 1 ...................................................................................... 6

4.4. Module 2 …………………………………………………………………………………..…....… 6

4.4.1. Functional Description .................................................................................................. 6

4.4.2. Modular Decomposition ................................................................................................ 6

4.4.3. Design Constraints ....................................................................................................... 6

4.4.4. Other Description of Module 2 ...................................................................................... 6

**Chapter 5: System Testing and Verification**

5.1. Testing Setup ……………................................................................................................... 7

5.2. Testing Plan and Strategy................................................................................................... 7

5.2.1. Module Testing ............................................................................................................. 7

5.2.2. Integration Testing ........................................................................................................ 7

5.3. Test Schedule.......................................................................................................................7

5.4. Comparative Results to previous work……………………………….......................................7

**Chapter 6: Conclusions and Future Work**

6.1. Faced Challenges ................................................................................................................8

6.2. Gained Experience ............................................................................................................... 8

6.3. Conclusions ..............................................................................................................8

6.4. Future Work ........................................................................................................................ 8

References .................................................................................................................................. 9

Appendix A: Development Platforms and Tools ..…………………................ 10

Appendix B: Use Cases ............................................................................................................ 11

Appendix C: User Guide ............................................................................................................ 12

Appendix D: Code Documentation ............................................................................................ 13

Appendix E: Feasibility Study .................................................................................................... 14

**List of Figures**

Figure 1.1: Same caption as in the text ................................................................................. page

Figure 2.1: Same caption as in the text ................................................................................. page

**List of Tables**

Table 1.1: Same caption as in the text .................................................................................. page

Table 2.1: Same caption as in the text .................................................................................. page

**List of Abbreviation**

[The abbreviations should be put in an alphabetical order]

AI Artificial Intelligence

EA Evolutionary Algorithms

GA Genetic Algorithms

SA Simulated Annealing

VLSI Very Large Scale Integration

**List of Symbols**

[The symbols should be put in an alphabetical order. Greek symbols come first, followed by English symbols]

σ Noise standard deviation

B Buffer size

fop Operating frequency

**Contacts**

**Team Members**

|  |  |  |
| --- | --- | --- |
| Name | Email | Phone Number |
| Ahmed Salama Hamed | [asivo.ahmed@gmail.com](mailto:asivo.ahmed@gmail.com) | +2 01144697687 |
| Ahmed Maher | Demha.ahmed@yahoo.com | +2 01065725338 |
| Moamen Hassan Attia | moamenattia@outlook.com | +2 01200370750 |
| Mohamed Talaat Mohamed | [mohamedtalaat0111790@gmail.com](mailto:mohamedtalaat0111790@gmail.com) | +2 01115598525 |

**Supervisor**

|  |  |  |
| --- | --- | --- |
| Name | Email | Number |
| Dr. Magda Fayek | [magdafayek@gmail.com](mailto:magdafayek@gmail.com) | +2 01xxxxxxxxx |

This page is left intentionally empty

**Chapter 1: Introduction**

This chapter presents the importance of soccer video analysis and summarization. Soccer video matches always attract major sports audience. Recently, the amount of digitized video content has been increasing rapidly and watching a soccer match needs a lot of time, many TV fans of sport competitions prefer to watch a summary of soccer video matches. Presenting the problems that are facing the automatic soccer video summarization and the proposed solutions. Finally an overview about the organization of the thesis is shown at the end of the chapter.

* 1. **Motivation and Justification**

Automatic soccer video summarization systems extracts the most important events to produce general summaries for the important moments in which soccer viewers may be interested. Researchers have proposed many techniques to take full advantage of the fact that sport videos have typical and predictable structure, recurrent events, consistent features, and fixed number of camera views and angles.

The multimedia analysis tool, which could automatically parse soccer video and output required video clips or the most interesting events such as goals, dangerous attacks and free kicks, fans could go though many more games without spending much time. This can certainly entertain these fans and in turn popularize the sport itself. So, soccer video analysis is absolutely necessary.

Event detection in soccer video is a high level analysis, which needs an effective description of soccer video information and approaches to bridge the gap between low level features (color, texture, shape and motion) and high level representation such as (shot types, shot length, and shot replays) However, research in this field is far from enough. Shot is commonly used as an intermediate representation, but the standard for soccer video analysis has yet to be reached and other high level representations should be explored. This work has been inspired by this motivation.

We have two propellants which motivated us to develop an automated system for soccer match summarization. First, most people cannot watch the whole matches which are played on same time within different time zones because of lack of time. Second, coaches need to view the highlighted events to truly developing plans and evaluate the team players. From this point we concluded the importance of our proposed program to put a solution for the mentioned problems.

* 1. **Project Objectives and Problem Definition**

Analyzing general sport games is still an open problem because of the variance and

diversity of different games and different sports. Some former researchers have proposed many highlight summarization methods both for general sports game and for a specific kind of sports game (2). detected the play and break event in sports videos to generate the summary. Some other researchers summarize sports videos using slow motion replays (3). On the other hand, another group of researchers turn to study specific sport games such as soccer, basketball or diving (4).

The summarization phase is an essential part in almost any sytem trying to retrieve the important parts in a video. This field is undergoing rapid change, as computers are now prevalent in virtually every application, from games for children through the most sophisticated planning tools for governments and multinational firms.

When we are talking about a soccer game, we can refer to it as a continuous sports which means that if there is an existence of replay during the match, it can be an indicator of the occurrence of important event such as (goal, penalty shot and red / yellow card). Therefore the summarization process which we aim for can be recognized by a combination of these events.

The input is a soccer video match needed to be summarized using a computer based application, our concern here is to extract the most exciting events in the soccer game such as (goal, attacks, and other events) using our proposed application then output those events into summarized video.

In this thesis, we are going to highlight the most important events such as (goals, attacks, and the other events) to save the viewer's time, and try to improve on the research that has benn done in the technology of computer based summarization of sports.

* 1. **Project Outcomes**

Proposed solutions have been presented to avoid most of the problems discussed in the previous section, these solutions are involved in each of the stages of the proposed system . In the preprocessing stage, reading all frames of a video is clearly an overhead on the sytem as most soccer matches are two hours long and also the differences between two consequtive frames doesn’t carry much information because very little has changed between the frames so frame skipping is done to ensure that there is no redundancy in the consequitve fremes and also make the system much faster.

shot boundary detection phase is proposed to segments the whole input video stream into small video shots by applying a combination of image processing techniques and histogram comparison techniques. In the shot classification phase , the system applies different algorithms, namely; Grass dominant ratio extraction, Face Detection ,Deep learning and image processing techniques to classify the shot into one of the following classes :wide – medium- close – close out – logo. The logo classification is used for replay detection which is an important factor in determining which of the shots are important and which are not.

In The event detection phase each is shot in sconsedered seperatly and compared with the previous shot to determine if a goal has happened between the two shots by comparing the scoreboard in the upper left corner. Two methods are proposed to detect a goal event, Optical character recognition (OCR) and structural similarity image index (SSIM).

Lastly, the summarization phase which takes into account the output of the previous phases to decide if he shot is important and shoul be included in the final video or should be discarded. For example, if a shot contains a goal then it is important or if a shot has a goal mouth and the audio level is high then it is important or it’s a replay shot and a replay of course is for something important. Then combining the important shots only in a video and that is the output of the system

* 1. **Document Organization**

In this section, you have to give the organization of the report and a quick description of the following chapters.

**Chapter 2: Market Visibility Study**

There are many platforms where a user can watch a summary of a soccer match such as “youtube”, “Filgoal” and “YallaKora”, but in these platforms the summaries are produced manulay, and even some online platforms provide only datasets for action spotting in soccer but do not provide a summarized video of these actions.

An automated system for summarizing may be available to corporation or clubs but there isn’t a product that is available for the mass public to use and that was an additional motivation for embarking on this project

**2.1. Targeted Customers**

As mentioned before such a system’s output is intended for the dedicated sports fans especially soccer fans who are not able to able to keep up with all the matches they want to whats so they resort to watching summaries of the matches they misses on such platforms such as youtube

**2.2. Market Survey**

In this section, list the competitive products to your work. Similar commercial tools/platforms should be mentioned and discussed. Write a subsection for everyone of them and explain its pros and cons in that subsection

**2.2.1. FilGoal and YallaKora**

These two platforms are very similar in the service they provide to the audience, they both act as online journals that provides audience by soccer news all over the world every day about all leagues and chamionships, Altough they provide news and information about other sports but their main focus is on soccer, for every match, they provide the events happeninig in the match minute by minute, they provide a video for each scored goal, finally, after the match ends, they provide a video of all of the scored goals.

The main advavntages of such platforms are:

* Provide soccer audience by the updated news about soccer all over the world
* Make their audience feel as if they were watching the match live by sending updates by events happening each minute in the match.
* Provide a video containing goals scored after each match.

The main disadvantages are:

* Do not provide a video summary of the highlights of the match (only goals).
* All of the services provided to the audience are done manually, which consumes a lot of money and time.

**2.2.2. Youtube Channels**

Many Youtube channels that are specialized in soccer provide a neat and clean summary after each match containing all the highlights and goals in the match.

These channels are the main competitive to the proposed project as they provide exactly the same service and for free.

The main advantages are:

* Provides a neat and clean match summary
* Free

The main disadvantages are:

* Done manually, which consumes time and money.
* Most of the channels do not have the rights to share the videos online, which causes them to get copy right strikes.

**2.3**. **Business** **Case** **and** **Financial** **Analysis**

In this section you describe the success of establishing a company to sell your product (or service)

Two Aspects must be addressed

Business Case: Based on Market survey above you should anticipate how many products you will sell over the next 5 years and how will you set your price to counter the competition.

**Business Case:**

* The proposed service will be provided to the audience as a website serving the summarized video of matches, user can search for any match he/she wants and watch its summary.
* The service will be available for free to be a strong competitor to the service provided by the “Youtube” channels, and will be dedicated to soccer video summarization to be a strong competitor to “FilGoal” and “YallaKora”.
* The main source of revenue will be based on online advertisement that will be presented on the website, there are many ways to make money with a website such as, “Affiliate Marketing”, “Pay Per Click” advertising, “Sell ad space”,“Sell sponsored posts”,etc.
* Over the next 5 years the revenues will mainly depend on the number of visitors of the website per month, the more the website visitors the more the advertisements in the website, and this means more money.

Financial Analysis: Based on the business case we must anticipate

1. The Capex (Capital Expenditure): These are one-time spending that you pay for development and buying things for the company
2. The Opex (Operational Expenses): These are recurring payments for salaries and marketing and … etc.

**Financial Analysis:**

* **Capex** (Capital Expenditure):
* Starting by 4 developers in the company, 4 laptops = 15,000 EGP/item,

4 desks = 2,500 EGP/item and 4 chairs = 1,500 EGP/item are needed.

* One air conditioner in the development room and another in the meeting room

= 10,000 EGP/item

* **Opex** (Operational Expenses):
  + Renting an apartment with two rooms (development room and meeting room), a kitchen and a bathroom = 7,000 EGP/month
* Salaries = 8,000 EGP/month for each developer.
* Marketing (Youtube advertisement) = 5$/day = 2400 EGP/month
* Renting servers (storage: 100 GB, RAM: 4 GB, speed = 100 Mbps)

= 550 EGP/month

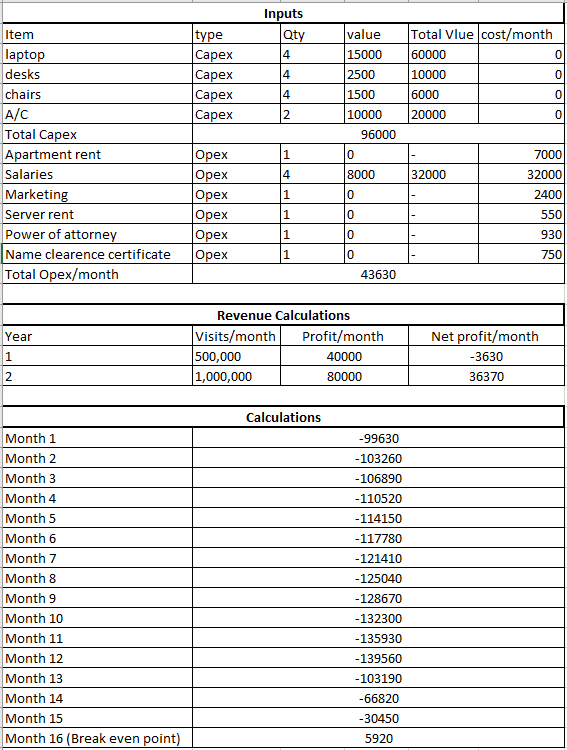
* Power of attorney = 930 EGP/month
* Name clearance certificate = 750 EGP/month
* **Revenue**
  + Will be based on online advertisement using “Sell ad space” method and will be calculated by how many visitors the website gets.
  + Typically it’s quoted as a dollar amount per one thousand impressions (or CPM). On average it is $5 = 80 EGP CPM. If the website gets 100,000 visits a month, that ad price translates into $500 bucks = 8000 EGP.
  + The good thing about this approach is that if the site gets a ton of traffic from different sources, a simple banner ad pricing can go up to as high as $5000 = 80,000 EGP per month! The obvious downside is that if the site doesn’t get a lot of traffic, you can’t expect to earn much either.
  + Assuming that the website will get 500,000 visits/month = 40,000 EGP/month in the first year and the number will double every year.

Then you create what we call a cash flow table (on an excel sheet). In this sheet you put down your monthly capex and opex on a set of rows and your reveneus (money you get back from selling product/services) on another set of rows.

The difference between both sums is your profit before tax.

It is likely that this difference is negative at beginning until your sales increase and counter the expenses.

From this cash flow analysis you find the date of the break even point wbich is the date at which all the money you get back equals the money you spent. From that date onward you will be making true profit ☺.

****

**Chapter 3: Literature Survey**

In this chapter we review the background on how a generic video is structured to be able to properly understand the processing that happens in the following chapters and also discuss

**3.1. Background on soccer analysis**

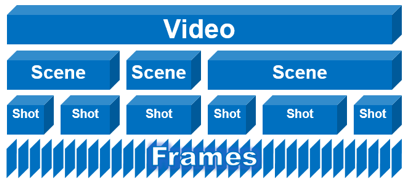
There are many works for detection of camera breaks in the past few years. Some typcal methods for the detection of camera breaks could be found in (8)(9)(10)(11). Recent published papers for shot change detection could be found In(12)(13)(14)(15)(16)(17).

Most work has been focusing on pixel difference, intensity statistics comparison, histogram distance, edge difference, and motion information. Among these methods, histogram basal ones have been consistently reliable, while DCT coefficient based ones give the lowest precision. Motion information based methods are somewhere in between. Some work for performance evaluation of shot detection could be found in (19)(20).

Some work has been done on detecting these special effects. Related works can be found in (21)(22)(23)(24). In a recent review paper, Lienhart (17) compares four major shot boundary detection algorithms, which include fade and dissolve detection. Extensive experimental results also favor the color histogram based method (25) for shot boundary detection, instead of the computationally expensive edge change ratio method (26).

**3.2. Background on video structure**

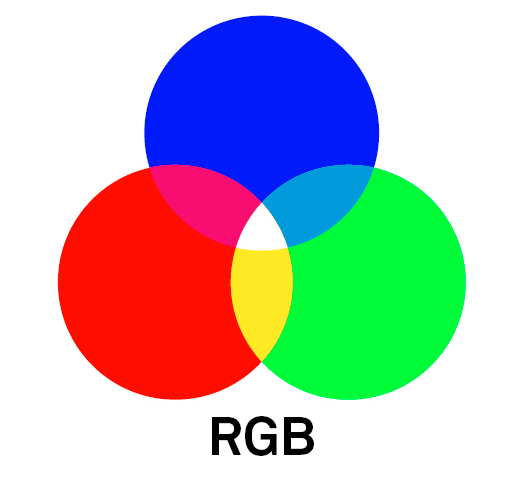
To probely understand the phases of the sytem, first we must understand a Video structure. Video data are typically organized in a typical hierarchical structure as shown in Figure, some elementary units such as scenes, shots, frames, key frame and objects are generated. A successful structure parsing is an important step in video classification and summarization. In the past, many works have been done in video structure parsing, especially in shot detection, motion analysis and video segmentation.



As discussed above, video data are structured into many shot units. Shot changes should be detected before dividing video data into shot units. shot change can be viewed as detection of a camera break. Normally, there are three major editing types of camera breaks: cut, wipe and dissolve. A cut is an immediate change from a shot to another shot; a wipe is a change where first frame of a shot is replaced with last frame of another shot gradually; a dissolve is a change where one shot gradually appears (fade-in) and another shot slowly disappears (fade-out). A cut can be detected by comparing two adjacent frames. While wipe and dissolve are difficult to detect since they are change gradually. The transition between shots usually corresponds to a change of subject, scene, camera angle, or view. Therefore, it is very natural to use shots as the unit for video analysis.

**3.3. Background on color spaces**

Various color spaces exist because they present color information in ways that make certain calculations more convenient or because they provide a way to identify colors that is more intuitive. For example, the RGB color space defines a color as the percentages of red, green, and blue hues mixed together. Other color models describe colors by their hue (shade of color), saturation (amount of gray or pure color), and luminance (intensity, or overall brightness).

The RGB color space represents images as an m-by-n-by-3 numeric array whose elements specify the intensity values of the red, green, and blue color channels. The range of numeric values depends on the data type of the image but normaly it’s in the range 0-255

**3.4. Comparative Study of Previous Work**

Y.H. Gong et al. in (40) proposed a system that can automatically parse soccer video programs using domain knowledge. The parsing process was mainly built upon line mark recognition and motion detection. They categorized the position of the play into several predefined classes by recognizing the compound line pattern with signature method. The motion vectors field is used to infer the play positions for those scenes without line marks. Despite the strong semantic indexes from the categorization of play positions, they have to address these two problems:

• How to identify different camera angle and shooting scale

• How to determine reasonable segment for processing.

D. Yow et al. in (41) presents techniques to automatically detect and extract the soccer highlights by analyzing the image contents, and to present these shots of action by the panoramic reconstruction of selected events. The analysis include the recognition of prominent features of the game, tracking of ball, camera movement compensation for effective recognition, and construction of the panoramic views. The authors pointed out a direction for application of soccer video summarization.

Utsumi et al. in (43) proposed a novel object detecting and tracking method in order to detect and track objects necessary to describe contents of a soccer game.

A. Ekin et al. in (2) presented an automatic and computationally efficient framework for analysis and summarization of soccer videos using cinematic and object based features. In this paper, algorithms of dominant color region detection are robust

A. Ekin et al. in (2) presented a fully automatic and computationally efficient framework for analysis and summarization of soccer videos using cinematic and object based features.

1..Y. Duan et al. in (45) presented a unified framework for semantic shot classification in sports video.

**3.4. Implemented Approach**

The Proposed system is presented to avoid most of the problems discussed in the previous sections, these solutions are involved in each of the stages of the proposed system . In the preprocessing stage, reading all frames of a video is clearly an overhead on the sytem as most soccer matches are two hours long and also the differences between two consequtive frames doesn’t carry much information because very little has changed between the frames so frame skipping is done to ensure that there is no redundancy in the consequitve fremes and also make the system much faster.

shot boundary detection phase is proposed to segments the whole input video stream into small video shots by applying a combination of image processing techniques and histogram comparison techniques. In the shot classification phase , the system applies different algorithms, namely; Grass dominant ratio extraction, Face Detection ,Deep learning and image processing techniques to classify the shot into one of the following classes :wide – medium- close – close out – logo. The logo classification is used for replay detection which is an important factor in determining which of the shots are important and which are not.

In The event detection phase each is shot in sconsedered seperatly and compared with the previous shot to determine if a goal has happened between the two shots by comparing the scoreboard in the upper left corner. Two methods are proposed to detect a goal event, Optical character recognition (OCR) and structural similarity image index (SSIM).

Lastly, the summarization phase which takes into account the output of the previous phases to decide if he shot is important and shoul be included in the final video or should be discarded. For example, if a shot contains a goal then it is important or if a shot has a goal mouth and the audio level is high then it is important or it’s a replay shot and a replay of course is for something important. Then combining the important shots only in a video and that is the output of the system

**Chapter 4: System Design and Architecture**

**4.1. Overview and Assumptions**

The proposed system is composed of five phases. Namely; in the pre-processing phase, the system reads the video stream into patches of frames. Then, in the shot Boundary phase is applied on every patch of frames to segment the video into seprate shots i.e. finding the transition from a camera into another. Then each shot is processed seperatly in the following three phases, the shot classification phase in concerned with classifiying each shot for example a wide shot is when the whole pitch is visible, A close shot is when the players faces are visible etc. then the event detection phase in which each shot is analyside to determine if a goal happens in it or not or if it the audio level in this shot is high with respect to the whole video or if the goal mouth appears in the shot . And lastly, the summarization phase in which the output of all previous phases is analysized to determine which shot is important and should include in the final video and which shots to discard

In this section, introduce how you design you system and develop its underlying architecture. Any employed assumptions should be clearly enumerated and justified.

**4.2. System Architecture**

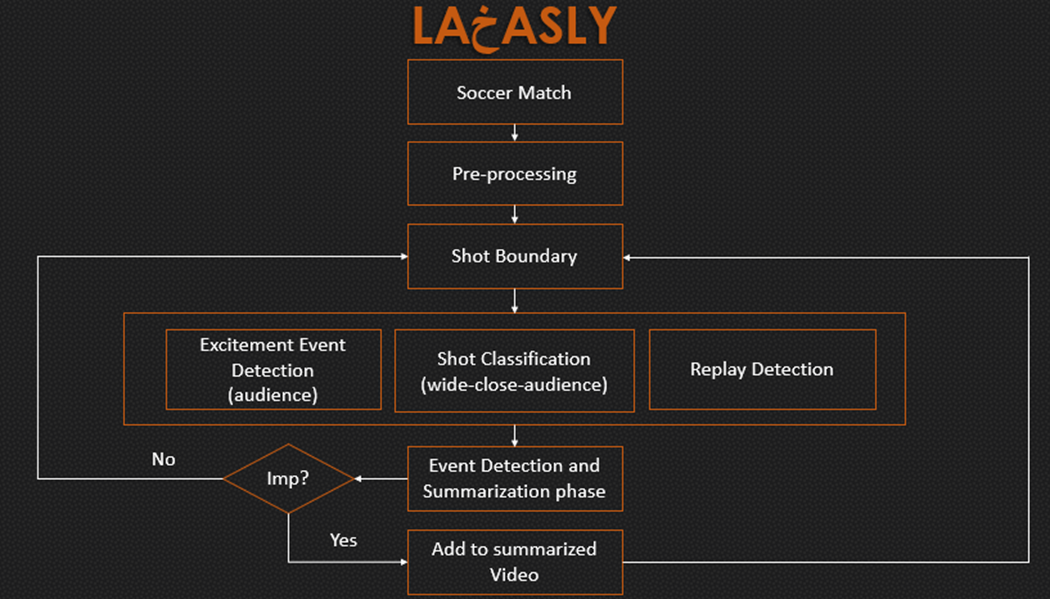
The Proposed system is presented to avoid most of the problems discussed in the previous sections, these solutions are involved in each of the stages of the proposed system . In the preprocessing stage, reading all frames of a video is clearly an overhead on the sytem as most soccer matches are two hours long and also the differences between two consequtive frames doesn’t carry much information because very little has changed between the frames so frame skipping is done to ensure that there is no redundancy in the consequitve fremes and also make the system much faster.

shot boundary detection phase is proposed to segments the whole input video stream into small video shots by applying a combination of image processing techniques and histogram comparison techniques. In the shot classification phase , the system applies different algorithms, namely; Grass dominant ratio extraction, Face Detection ,Deep learning and image processing techniques to classify the shot into one of the following classes :wide – medium- close – close out – logo. The logo classification is used for replay detection which is an important factor in determining which of the shots are important and which are not.

In The event detection phase each is shot in sconsedered seperatly and compared with the previous shot to determine if a goal has happened between the two shots by comparing the scoreboard in the upper left corner. Two methods are proposed to detect a goal event, Optical character recognition (OCR) and structural similarity image index (SSIM).

Lastly, the summarization phase which takes into account the output of the previous phases to decide if he shot is important and shoul be included in the final video or should be discarded. For example, if a shot contains a goal then it is important or if a shot has a goal mouth and the audio level is high then it is important or it’s a replay shot and a replay of course is for something important. Then combining the important shots only in a video and that is the output of the system

**4.2.1. Block Diagram**



**4.3. pre-processing phase**

As the first phase of the system it’s functions are simple which are to read the video to be summarized form storage, skip unnecessary and redundant frames to avoid any unnecessary calculations and store the frames in a proper data structure for subsequent phases to process.

As mentioned before, reading all frames of a video is clearly an overhead on the sytem as most soccer matches are two hours long and also the differences between two consequtive frames doesn’t carry much information because very little has changed between the frames so frame skipping is done to ensure that there is no redundancy in the consequitve fremes and also make the system much faster.

The frame skipping process consists of reading a frame then ignoring k sunsequent frames then appending the k+1 frame. The value of k is best within the range (5,10) since less than 5 there will be useless frames and larger than 10 there will be missing important information.

Reading a 2-hour video into a computer’s memory especially if the video is of higher quality would require too much resources, the solution for this is to divide the input frames into patches each is roughly 2000 frames which is good for an eight giga byte memory.

The processing of the patch through all phases of the system exept the summarization phase then repeating the pre-process phase to obtain the next patch ans that whole process continues until the end of the video then the summarization phase can start

**4.3.1. Functional Description**

**4.3.2. Modular Decomposition**

Explain the modular decomposition of the coarse module into smaller fine ones

**4.3.3. Design Constraints**

Explain the constraints that affect the design of the module

**4.3.4. Other Description of Module 1**

Give any other necessary discussion of the module to ensure that it is clearly described.

**4.4. Shot Boundary**

before defining the term of shot which is used frequently in this chapter, we are going to introduce some basic concepts and techniques applied in shot boundary detection.

First, What is a shot? It is a view that comes from a certain camera and is different from a view that comes from another camera. in a soccer match, views come from different cameras positioned at different locations along the pitch. You (as a human) can realize when a transition between different cameras (a cut) happen but how would a computer know given the same information

We have two types of transitions in a cocker match

1) instant (cut) transition

2) gradual transition.

A gradual transition is When there is a special editing effects during the match to illustrate the transition such as a logo of the competition in which the match takes place.



Figure 5 example of an instant cut



Figure 6 example of a gradual cut

After discussing the previous basic concepts that are very important , this section presents the hierarchal steps of the shot-boundary detection algorithms simply by following the flowchart in Figure (4.4).

We use color histogram of two consecutive frames as the main identifier for the shot boundary detection phase since color histograms are robust to moderate object and camera motions, we represent a frame by its color histogram, which is defined in the RGB space i.e eah bin in the histogram corresponds to a combination of values (r,g,b). we used 64 bins for each channel which means that now we get a 64x64x64 array, where each element corresponds to number of pixels with a particular (r,g,b) values. For example, histogram[50,100,150] denotes number of pixels who have the values B=50, G=100, R=150.

The similarity between the two histograms is measured by histogram intersection and histogram correlation.

Comparing histogram is a powerfull process to know if two images are different or not and measure the similarity. there many method to compare two histograms but the ones considered are discussed below

• Intersection.

**Intersection of two histograms h1, h2**

d(H_1,H_2) =  \sum _I  \min (H_1(I), H_2(I))

This method simply compares, for each bin, the two values in each histogram, and keeps the minimum one. The similarity measure is then simply the sum of these minimum values. Consequently, two images having histograms with no colors in common would get an intersection value of 0, while two identical histograms would get a value equal to the total number of pixels.  
• Correlation.

**Correlation of tow histograms h1, h2**

d(H_1,H_2) =  \frac{\sum_I (H_1(I) - \bar{H_1}) (H_2(I) - \bar{H_2})}{\sqrt{\sum_I(H_1(I) - \bar{H_1})^2 \sum_I(H_2(I) - \bar{H_2})^2}}

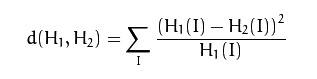
where

\bar{H_k} =  \frac{1}{N} \sum _J H_k(J)

is based on the normalized cross-correlation operator used in signal processing to measure the similarity between two signals, cross-correlation is a [measure of similarity](https://en.wikipedia.org/wiki/Similarity_measure) of two series as a function of the displacement of one relative to the other. This is also known as a sliding [dot product](https://en.wikipedia.org/wiki/Dot_product) or sliding inner-product.

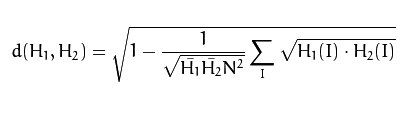
For the Correlation and Intersection methods, the higher the metric, the more accurate the match

* Chi-Square



It’s a simple method sums the normalized square difference between the bins of the histogram

* Bhattacharyya distance



Bhattacharyya distance is used in statistics to estimate the similarity between two probabilistic distributions.

the Chi-Square and Bhattacharyya distance methods, the higher the metric, the lower the match between the two histograms

through trial and error it was shown that that the intersection and correlation maethods are the most accurate therefore reliable as they profuced consistant outcomes in the instant cut as well as the gradul cut on the other hand Chi-Square and Bhattacharyya distance had low accuracy in classifying the gradual cut. So, the proposed algorithm is dependant on the intersection and correlation methods.

The proposed shot-boundary detection algorithm is able to detect both instant and gradual transitions with a high detection and low false alarm rates. The detection of gradual transitions in sports video is particularly difficult because of the high color correlation between two shots so Instead of computing the histograms for every consecutive frame pair, the comparisons are performed between frame i and frame i + placement k as demonstrated before.

The two frames below are obviously consecutive frames from the same shot i.e. there is no cut.

that should be reflected in the values of histogram intersection and correlation



*Histogram Intersection = 9.04120311407678*

*Histogram correlation = 9.971235345958373*

The values are higher than the thresholds mentioned in the above algorithm (the thresholds are based on trial and error) so no cut

Here is another example to fully explain how the shot boundary algorithm works.

These are also consecutive frames but there is a transition (cut) between them



*Histogram intersection = 0.24404063194742776*

*Histogram correlation = 0.002219072222348816*

Both values are smaller than the thresholds so we perform step 4 of the algorithm in which dividing the two frames into blocks, each block 150px x 150px and calculate intersection and correlation between individual blocks then, according to the values of intersection and correlation of the blocks to consider whether that block is considered full changed or half changed or not changed at all.

Then, step 9 in which we count how many blocks have changed and calculate the percentage with respect to all block

In the above two frames the percentage of changed blocks is 100% i.e. all frame blocks have has changed so, there is a cut.

Another example to full demonstrate the importance of having two methods.

The following two frames are consecutive but in the midst of a logo transition (gradual transition)



*Histogram intersection = 2.1293090697099615*

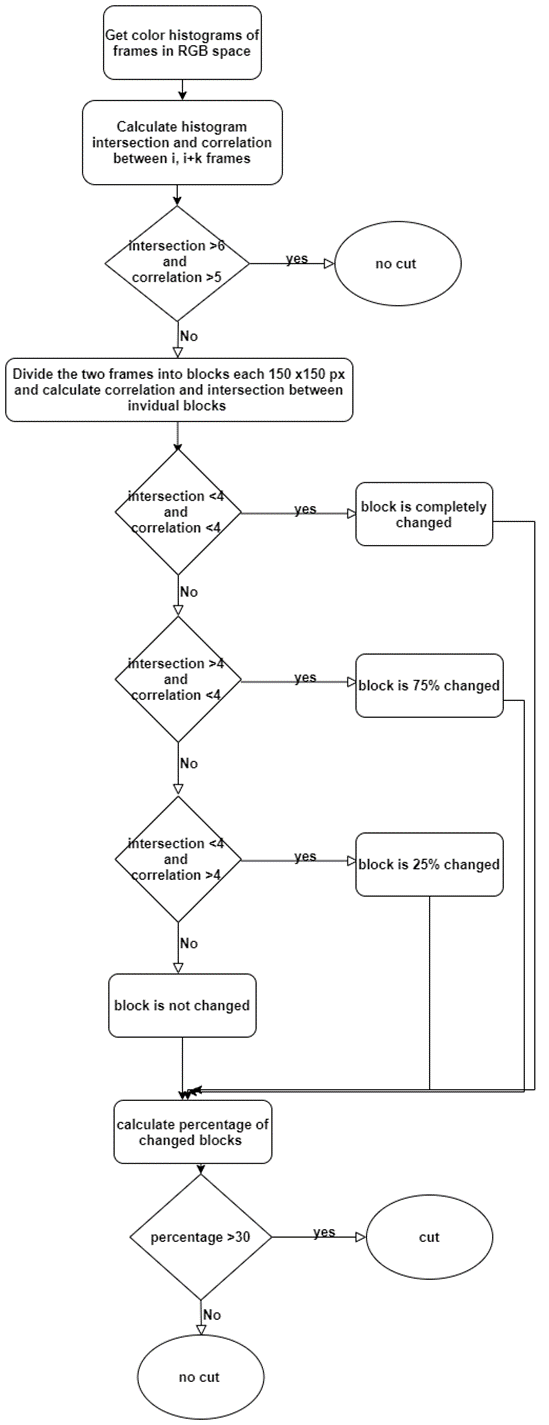
*Histogram correlation = 9.844120057485238*

Here the two methods disagree, the intersection indicates that there are minor similarities in the two frames but the correlation indicates that that’s they are variations of each other i.e. the histogram is a shifted version of the other histogram for example.

So, we resort to the blocks process once again to determine if there is a cut or not. The percentage of changed block is 37% which makes sense because there are areas that didn’t change but according to the threshold this is still considered a cut and that is logical because it’s a logo transition.

**Algorithm:**

* Get color histograms of the two frames in RGB space
* Calculate histogram intersection and correlation between the two histograms
* If intersection >6 and correlation >5 then no cut
* Divide the two frames into blocks, each block 150px x 150px and calculate intersection and correlation between individual blocks.
* If intersection <4 and correlation <4 then block is 100% changed
* If intersection >4 and correlation <4 then block is 75% changed
* If intersection <4 and correlation >4 then block is 25% changed
* If none of the above conditions are met then the block is not changed
* Count changed block and calculate percentage of changed blocks. 10- If percentage of changed blocks > 30% then Cut else no cut



**4.4.1. Functional Description**

Explain the functional description of the module

**4.4.2. Modular Decomposition**

Explain the modular decomposition of the coarse module into smaller fine ones

**4.4.3. Design Constraints**

Explain the constraints that affect the design of the module

**4.4.4. Other Description of Module 2**

Give any other necessary discussion of the module to ensure that it is clearly described.

**4.5. Exitment Event detection**

**4.5.1 Audio Processing**

Loudness, silence and pitch generated by the commentator and/or crowd are effective measurements for detecting excitement. The volume level is the most frequently used and simplest audio features as an indication of the loudness of the sound, so in this module we get the video times in seconds where the volume is high which is an indication of an important event in the match.

He below algorithm samples an audio level that is the average value from every 10 seceonds in the match then calculates the difference between consiqutive averages i.e. consequitve 10 seconds of the match.

If the outcome of the diffrencing operation is positive +ve that mean the volume is increasing so if multiple increasing instances are consequtive that is an indication of a perion in the match where the audio levels are rising.

We find that by calculating the 90% of the values in the averaged volumes.

for example: averaged volumes = [1,2,3,4,5,6,7,8,9,10,11,…,97,98,99,100]

so, having the averaged array, we take the largest 10% of it

so, values > 90% are [91, 92,…,99,100] i.e. the larger 10% of the array.

Algorithm:

1. Read video clip:
2. Extract audio from the video clip:
3. Get average volume of each 10 seconds
4. Get the difference between every two averages then detect the increases and decreases in volume:
5. Determine peaks indices of volumes:
6. Get peaks volumes:
7. Get Times of peaks having volume level > 90%

**Flow Chart**

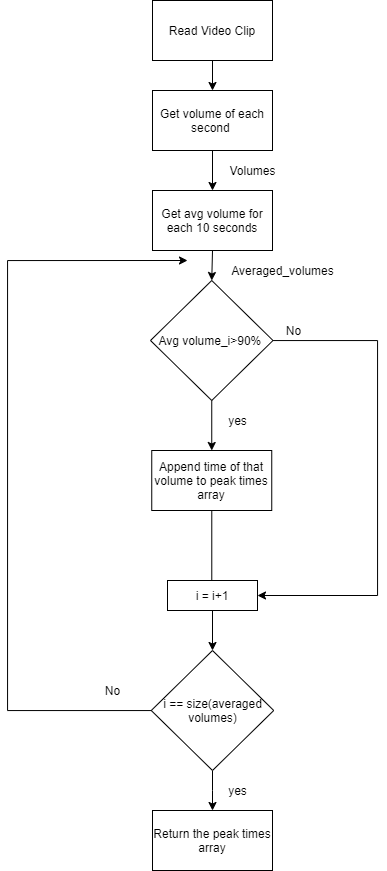


Figure 5 Flow chart for the audio processin

# 4.5.2. Goal Mouth Detection

## Introduction

For soccer video, the goal-mouth scenarios can be selected as the highlighted candidates, for the reason that most of the exciting events occurs in the goal mouth area such as the goal, shooting, penalty, direct free kick, etc.

On the other hand, the non-goal-mouth scenarios often consist of the dull passes in the midfield, defense and offense or some other shots to the audiences or coaches, etc., which are not considered as exciting as the former.



Figure 6 Examples of important events containing goal post

## Goal-mouth detection algorithm

Goal-mouth can be presented in wide shots and medium shots for each one there is a different image processing approaches to detect if the shot contains goal-mouth or not due to different camera angle, point of view, etc.

For wide shots, The appearance of two or three parallel field lines in a wide view can be used to indicate the occurrence of the goal-mouth. The appearance of goal-mouth and parallel field lines are highly co-related. This information of parallel lines which indicates the penalty box is very useful for goal-mouth detection. The information of the parallel lines is more reliable than the information of the goalpost here in the wide shot, since the goalposts detection may fail due to the cluttered background pixels and the environment interference. Here, we use canny edge detector with minimum value and maximum value threshold of 60 and 120 and Hough Transform to detect the lines of minimum length of 80 giving the endpoints of each line. The parallel lines are detected by checking if the difference between their angles is less tan a tolerance threshold, and their slope angles range from 140 to 170 or range from 10 to 40. When parallel lines tilt to left, it implies a right goal, otherwise a left goal [].

For medium shots, RGB image is converted into gray colour image in order to do further image processing operations. Noise in the image has been reduced using Gaussian low pass filter. Canny edge detection algorithm with minimum value threshold of 50 and maximum value threshold of 100 is used to detect the edges in the image for identifying the goal post region. The goal post region is obtained by applying masking on edge detected image, the masking is derived by mapping the edge detected image with the dominant green colour image. The horizontal and vertical lines parallel lines of minimum length of 60 are identified to find the goal-mouth in the region of interest using Hough transform [].

***Summarized steps:***

* Take input image frame and its type(wide or medium).
* Convert RGB image to gray colour.
* Remove noise using Gaussian filter.

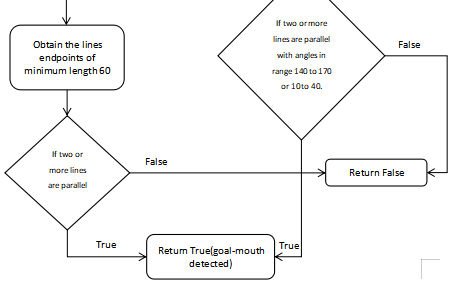
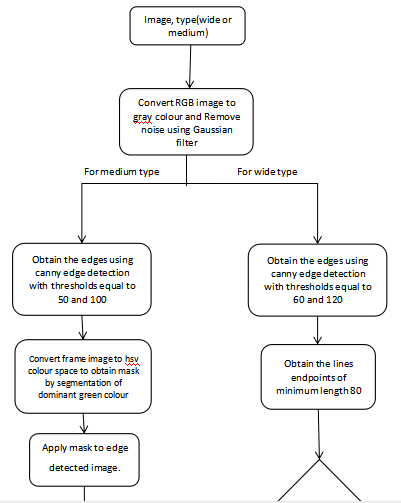
For wide type:

* Get the edges using canny edge detection with maxVal and minVal equal to 60 and 120 respectively.
* Obtain the lines endpoints of minimum length 80 from the edge detected image using Hough transform.
* If there is two or more lines parallel with angles in range 140 to 170 or 10 to 40 then return true and goal-mouth is detected.

For medium type:

* Get the edges using canny edge detection with maxVal and minVal equal to 50 and 100 respectively.
* Convert frame image to hsv colour space to obtain mask by segmentation of dominant green colour.
* Apply mask to edge detected image.
* Obtain the lines endpoints of minimum length 60 from the edge detected image using Hough transform.
* If there is two or more lines parallel then return true and goal-mouth is detected.

**Flow Chart**



# Excitement Event Detection (Goal detection)

## Introduction

The score board is a caption region (usually at the top left) distinguished from the surrounding region, which provides the information about the score of the game and the match time. The score board dimensions and position is always constant in a certain league so detection of the score board itself is not needed



Figure 8 Examples of Scoreboards

## Goal detection using OCR

The goals are detected by the help of scoreboard with optical character recognition



Figure 9 Example Before and after a goal

**La5asly** uses **Tesseract** engine for **OCR**, it is free software, developed by **Google.**

### Algorithm

***1: while reading the input video.  
2: for each 5 seconds do.  
 2.1: apply the mask to get the scoreboard.  
 2.2: run the tesseract engine to check if the results changed or not.  
 2.3: if changed and stable:  
 2.3.1: save the results (error free).  
 2.4: if changed:  
 2.4.1: save the results locally and go to 2.***

We need tesseract OCR to get the number in different shots to see of the number changed

## Goal Detection with structural similarity image index (SSIM)

### Abstract

In this approach, Goals are detected with the structural similarity image index (**SSIM**), It is an enhanced way to detect if the image changed or not. It uses the difference between the scoreboard and get the mean, variance and illuminance between the two scoreboards and get if the scoreboard changes or not. If scoreboard changes then it is an event (goal or substitution).

### Flow Chart

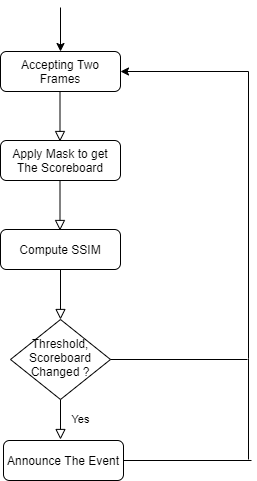


Figure 10 Flowchart for ssim

tesseract has an advantage over the ssim method, it returns only the goals.

we see the results as strings, so comparing is easy.

But for SSIM method, the results are percentages of change so it’s threshold dependent

And this is not good because in substitutions the scoreboard changed also to declare the substitution

example: [https://ibb.co/LNXP8v6](https://ibb.co/LNXP8v6?fbclid=IwAR2E-R86nooKpmq59BPLWlNTbrKGpFXpt92ogU0piWovq_LpgOfMz27zqkI)

**4.5.1. Functional Description**

Explain the functional description of the module

**4.5.2. Modular Decomposition**

Explain the modular decomposition of the coarse module into smaller fine ones

**4.5.3. Design Constraints**

Explain the constraints that affect the design of the module

**4.5.4. Other Description of Module 2**

Give any other necessary discussion of the module to ensure that it is clearly described.

**4.6. shot classification**

Production crews use different shot types in broadcasting a soccer match, which is  
be used for high-level video analysis in a particular domain.

Cinematographers classify a shot into one of four categories Wide, medium, close-up and audience (out-of-field) shot classes, the definitions of which are usually domain-dependent.

In the following, we define these four classes for sports videos:

1. **Wide Shot:** A long shot displays the global view of the field; wide shots almost always display some part of the stadium, which decreases the dominant colored pixel ratio.
2. **Medium (In-field) shot:** A medium shot, where a whole human body is usually visible, is a zoomed-in view of a specific part of the field.
3. **Close-up:** A close-up shot usually shows the above-waist view of a player or referee.
4. **Audience (Out-of-field) Shot:** The audience, coach, and other shots are denoted as out of field shots.

The sequence occurrence of a close-up shots and audience (out-of-field) indicates an important event such as (goal, goal attempts …etc.) during the match.

As we illustrated before in our meetings, after the classification we observe the sequence of shots to determine if an important event occurred or not.

For example, when a goal is scored in the wide shot the director of the match usually cuts to the player who scored the goal celebrating (i.e. close) then a shot of the audience (i.e. out of field) then a replay of the goal.

**

Figure 2 Sequence of shot to declare a goal

## Shot Classification Approaches

The above definitions consider shot-types as functions of field region. Because field  
region information is available after **Grass dominant ratio extraction** (discussed before in progress report-1) but using only the grass dominant ratio didn’t yield great results in shot classification.

In the following section we discuss four approaches to classification and comparing between them:

1. Shot classification using Grass dominant ratio extraction
2. Shot Classification using Face Detection
3. Shot classification using Deep learning
4. Shot classification using image processing techniques

### Shot classification using Grass dominant ratio extraction

The proposed algorithm is based on a specific threshold (range) for grass ratio (G); which was developed by observing different matches in many lighting and weather conditions to appropriately define a range that will cover the four shot types.

Classification of a shot into one of the discussed three classes is based on spatial features.  
Therefore, shot class can be determined from a single key frame or from a set of frames  
selected according to a certain criterion.

G is the grass ratio in the frame as mentioned in the first line of the above paragraph

For example: if the frame is in a wide shot that means that the whole filed is visible which means that G would be high (> ~70%)

Thinking intuitively, G would be almost zero for out-of-field frames, a low G value in a frame corresponds to a close-up, while high G value indicates that the frame is a long view, and in between, a medium view is selected.

Frame Type

***For each shot***

***Choose a set of key frames in the shot***

***Compute grass ratio G of these frames***

***Classify frames***

***Determine the majority type of the frames and assign it to the shot***

The set of the key frames is based on the length of the shot, if the shot is small (less than 50 frame) then the whole shot is included and if the shot is larger then sample 50 frames uniformly from the shot from start to end.

Choosing the set of frames also depends on the computation power running the code.

Due to the computational simplicity of the proposed algorithm, computing the grass ratio of all frames in the shot would not be a big overhead.

Although the accuracy of the above simple algorithm is sufficient but it does not meet the desired outcome.

The accuracy of the above algorithm is roughly 60% (i.e. it classifies 60% of the shots correctly)

### Shot classification using Face detection

In the observation phase mentioned above it can be said that the Wide shot does not contain any clear faces (large enough to be recognized by a human or a computer), in the medium shot a face can or cannot be recognized depending on the angle of the camera filming the shot, in the close shot a face can be clearly recognized as it covers a large area of the frame.

With that said, combining face detection model and the Ratio G to obtain better results in classification.

Frame type

***For each shot***

***Choose a set of key frames in the shot***

***Compute grass ratio G of these frames***

***Determine whether the frames contain face or not and if it does get the***

***Area of the bounding box around the face***

***Classify frames***

***Determine the majority type of the frames and assign it to the shot***

The obtained results are better than using grass ratio only but the computational time and complexity is much worse because the face detection model is very complex and time consuming.

This approach yielded good results (~70%) but we didn’t spend much time refining it because the face detection model is pretrained and we wanted our classification method to be implanted from scratch

### Shot Classification using deep learning

In this approach, Frames are Extracted from Matches and filtered (assigned labels) manually to construct a data set, Then A model **(CNN)** is used to train on the constructed data.

By ‘filtered’ we mean assigning different labels manually.

For example, we extract the frames from the match then separate the wide frames form the close from the medium.

It’s an exhaustive process to get labeled data, the model is trained on these labeled frames to be able to predict the type of a given input frame.

The model trained with 8,000 images (Total 40,000) per each class, 20 epochs. The input to model is the raw image 28x28x3.

Firstly, the model suffered from overfitting. We overcame this problem by adding dropout in layers with probability 0.6 and the results got better.

This approach is much better than the previous techniques and has a great classification accuracy 86%.

### Layers Explanation

**Convolution** is the first layer to extract features from an input image. Convolution preserves the relationship between pixels by learning image features using small squares of input data. It is a mathematical operation that takes two inputs such as image matrix and a filter or kernel. Convolution of an image with different filters can perform operations such as edge detection, blur and sharpen by applying filters.

**Pooling** layers section would reduce the number of parameters when the images are too large. Spatial pooling also called subsampling or down sampling which reduces the dimensionality of each map but retains important information.



Figure 3 CNN Model

### Shot Classification using image processing

In this approach, an image, after a dominant color mask is applied i.e. green is white and any other color is black, is divided into 3:5:3 grid.

The frame is divided into a 3:5:3 grid as illustrated in the figure below



Figure 4 frame after splitting into 3:5:3

The image classes are determined using the green pixels in 1, 2, 3 regions and the absolute difference between 1, 2 and 2, 3 to estimate is there a close shot or not.

The close out is pretty simple, if green ratio is tiny (less than 10%) then it is close-out.

But this approach has a good accuracy, actually near 60%, but compared to the above its subpar.

The accuracy is great for an image processing technique and no learning.

The advantages of this technique are that its computationally inexpensive (no models) and its implemented from scratch

**Comparison between shot classification approaches**

|  |  |  |  |
| --- | --- | --- | --- |
| Grass dominant color | Face detection | deep learning | image processing |
| **60%** | **70%** | **86%** | **60%** |

**4.6.1. Functional Description**

Explain the functional description of the module

**4.6.2. Modular Decomposition**

Explain the modular decomposition of the coarse module into smaller fine ones

**4.6.3. Design Constraints**

Explain the constraints that affect the design of the module

**4.6.4. Other Description of Module 2**

Give any other necessary discussion of the module to ensure that it is clearly described.

**4.7. Summarization and Classification phase**

The main purpose of this phase is to choose a number of important shots from the whole video shots, then these shots important shots are concatenated together to form the final summarized video.

After having the summarized video, we will classify the sequence of these important events.

The sequence of each important shots will be classified into 3 classes:

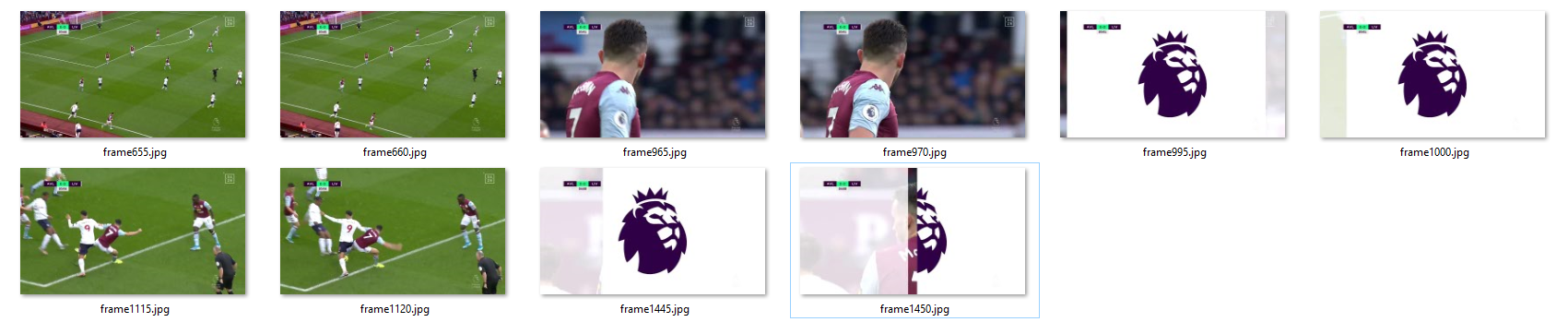
1. Goal Event
2. Attack Event
3. Other Event (fouls, offsides, substitution, etc.)

**Extracting Important Events**

Till this point we have some important information about each shot:

1. Shot frame number , shot start time, shot end time 🡪 output of **Shot Boundary**
2. Shot type 🡪 output of **Shot Classification**
3. Whether a goal is scord or not 🡪 output of **Goal detection**
4. Whether the shot contains a goal mouth or not 🡪 output of **Goal Mouth Detection**
5. Whether the shot contains high volume level 🡪 output of **Audio Processing**

Every replay in the match represents an important event such as, a replay for a scored goal or an attack or even a foul, so starting by detecting the replay shots ( two logo shots ), we can end up with a summarized video containing the most interesting events in the match, this process will start by detecting two consecutive logo shots, then because the first wide shot before the first logo is the shot that most likely contains the important event, we traverse backward till the the first wide shot before the first logo shot, finally, having theses set of important shots we append it to the output video. We start the same process again on the remaining shots.

****

**Methodology**

**Input:** shots array, each has **(** Shot frame number , shot start time, shot end time, Shot type,

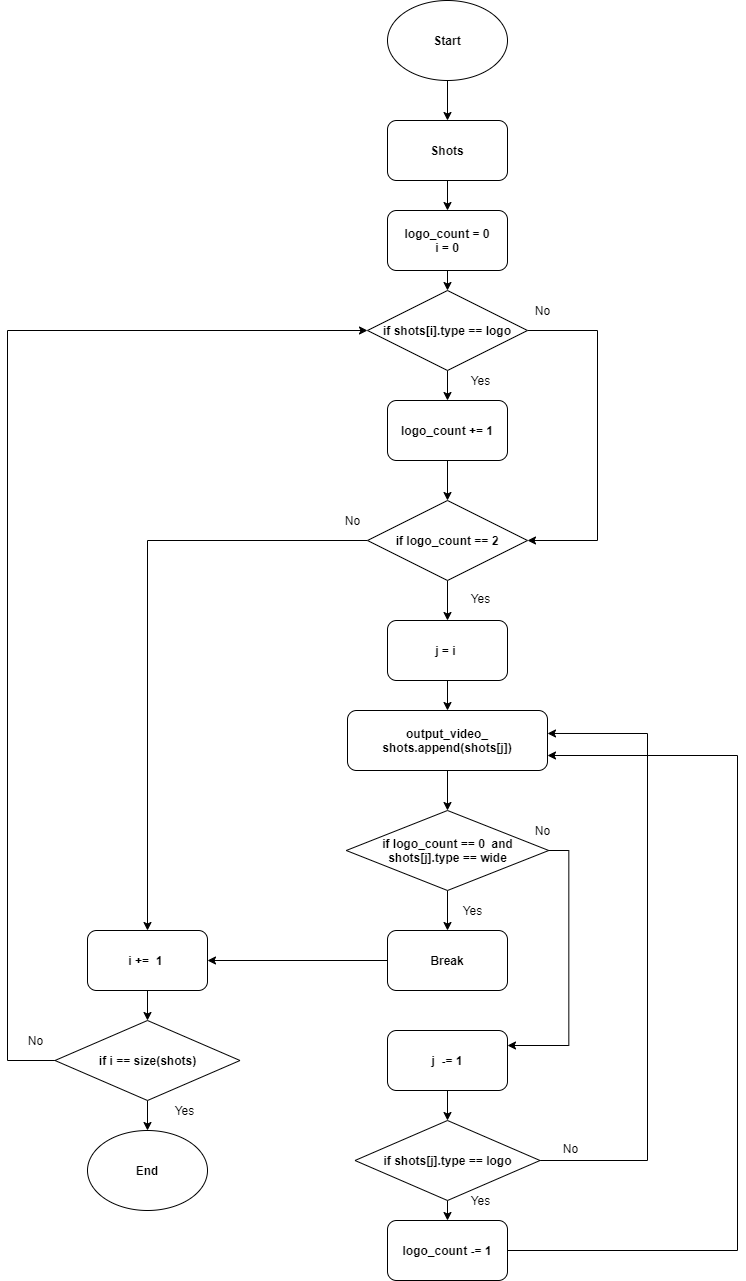
Whether a goal is scord or not, Whether the shot contains a goal mouth or not, Whether the shot contains high volume level)

**Output:** Summarized video containing only the important shots

**Algorithm:**

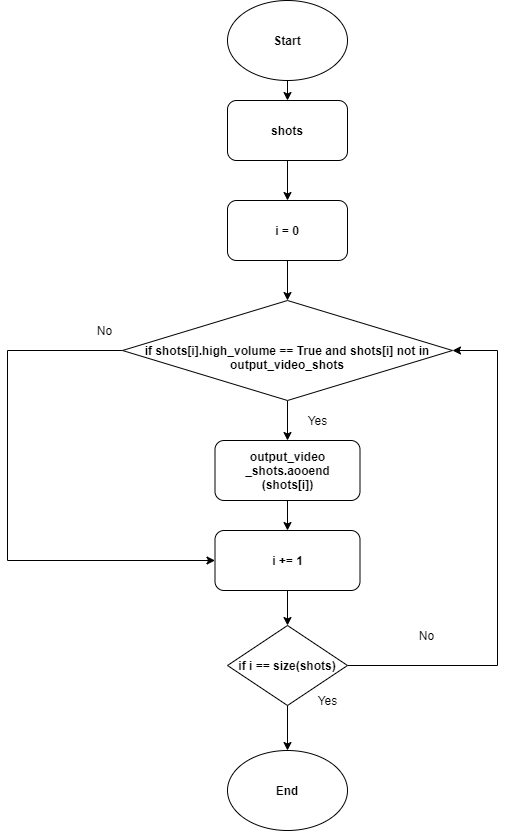
1. Logo count = 0
2. for each shot in the shots array
3. if shot type == logo:
   1. logo count +=1
4. if logo count == 2:
   1. append to the output video shots
   2. if shot type == wide and logo count == 0
      1. go to 3
   3. if the shot type == logo
      1. logo count -= 1
   4. move to the backward shot
   5. go to 4
5. repeat until all shots are processed
6. End up with the final summarized video

**Flow chart:**

****

After detecting the important events based on the replay detection we will try to find out important events in the match that are not replayed, we will depend on the output of **Audio Processing**, for each shot we know whether each shot contains high volume level or not, which is a strong evidence of whether a certain shot is important or not.

By appending the shots that contain high audio level but not included in a replay sequence, we end up having all the important events in the match.



**Shots Sequence Classification**

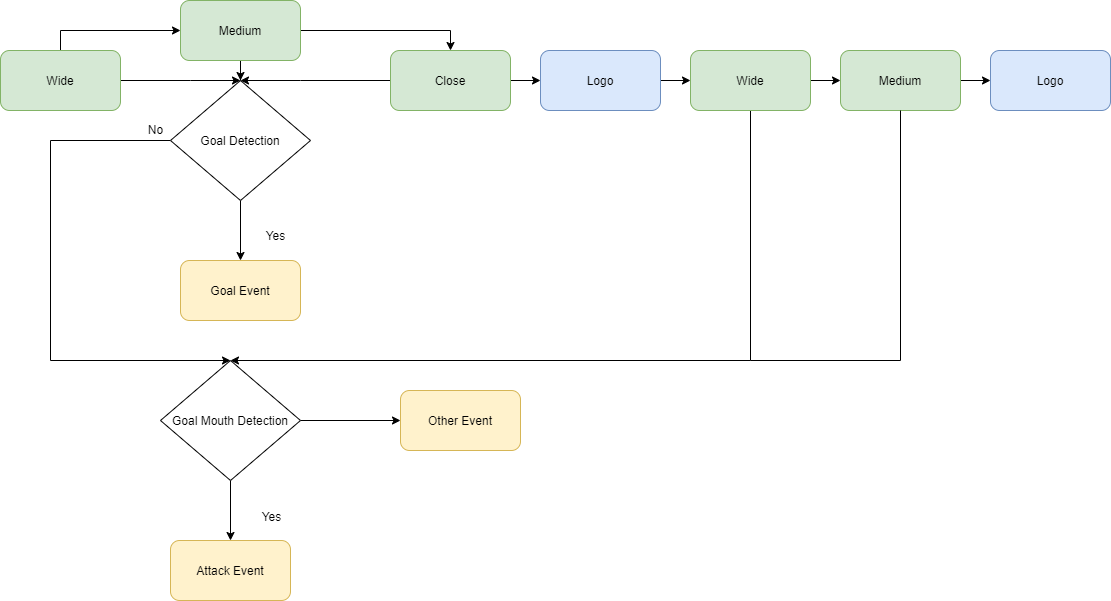
As we have to sets of important events:

1. Events come from replay detection
2. Events not replayed but have high volume level

We will classify each set individually, then collect the important events together and sort them based on the shot frame number, so that we have all the important shots from both sets in the right order.

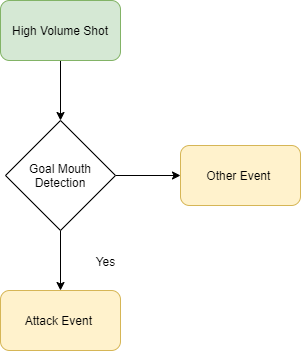
**Classifying shots that come from replays**

* When a goal is scored, the score in the score board is changed which can be detected by the **Goal Detection** phase and results in goal event.
* All of the attacks are most likely to happen near the goal mouth, so finding a goal mouth in a sequence of important shots and the score board is not changed, will result in attack event.
* Other important events such that fouls, offsides, substitutions, etc. that neither contain score boared change, nor happen near a goal mouth will result in other event.

****

**Classifying shots that are not replayed but contain high volume level**

In classifying this set of shots we are sure that we can not end up with a goal event as every scored goal in soccer matches is replayed for several times, so we have already obyained it in the replayed shots, fot this set of shots we will classify the sequence of the shots in only two classes, attack event and other event, our classification will be based on detecting a goal mouth in the shot, as said before, all attacks most likely to happen near a goal mouth so detecting a goal mouth is a good indicator that a certain shot is an attack, and if a goal mouth is not found, the shot will be classified as other.



**Chapter 5: Audio version**

In this chapter we presenta version of laخsly where the video summarization is based only on the audio feature i.e. and the shot boundary process i.e. when the audio level is high that gives an indication of an important event in the football match so we extract the shot in which that important event happended

**5.1. Reason to develop this version**

the initial results of the main version showed that theaudio feature is contributing significantly in the summarized final video so making a version with audio is worth the effort and would be much faster than the original version, with that said it’s not without its drawbacks

**5.2. Advantages**:

* Very less computations compared to the main version.
* Depends only on the Audio module and the shot boundary module but the main version has more modules to take into considerations.
* Much faster runtime

5.3. **Disadvantages**:

* The summarized output of this version is relatively longer in minutes than the main version as in many matches the audiences sometimes have loud voice while cheering even if it’s not really an important event in the game.
* The summarized video can get longer depending on audience and commentator attitude in every match.
* In rare situation, some important events are not included because the audio level is not larger than the audio threshold

5.4. **Used Modules Details**

The used moduls in this version of laخsly are Audioand Shot Boundary

1. **Audio module Input:** video clip

**Output:** times in video having volume level > 90% of the video volumesAlgorithm:

* Read video clip:
* Extract audio from the video clip:
* Get average volume of each 10 seconds
* Get the difference between every two averages then detect the increases and decreases in volume:
* Determine peaks indices of volumes:
* Get peaks volumes:
* Get Times of peaks having volume level > 90% [largest 10%]

1. **Shot Boundary module Input:** two frames **Output:** cur or no cut

**Algorithm:**

* Get color histograms of the two frames in RGB space
* Calculate histogram intersection and correlation between the two histograms
* If intersection >6 and correlation >5 then no cut
* Divide the two frames into blocks, each block 150px x 150px and calculate intersection and correlation between individual blocks.
* If intersection <4 and correlation <4 then block is 100% changed
* If intersection >4 and correlation <4 then block is 75% changed
* If intersection <4 and correlation >4 then block is 25% changed
* If none of the above conditions are met then the block is not changed
* Count changed block and calculate percentage of changed blocks. 10- If percentage of changed blocks > 30% then Cut else no cut

**Input:** Video Clip

**Output:** Summarized video clip

**Algorithm:**

1- Read video clip

2- Get Peaks times using Audio module, peaks [i…n]

3- For each peak time peaks[i], get its frame number

4- For each frame get its shot using the shot boundary module.

1- Starting from this frame explore frames descending tell finding a shot cut

= shot start

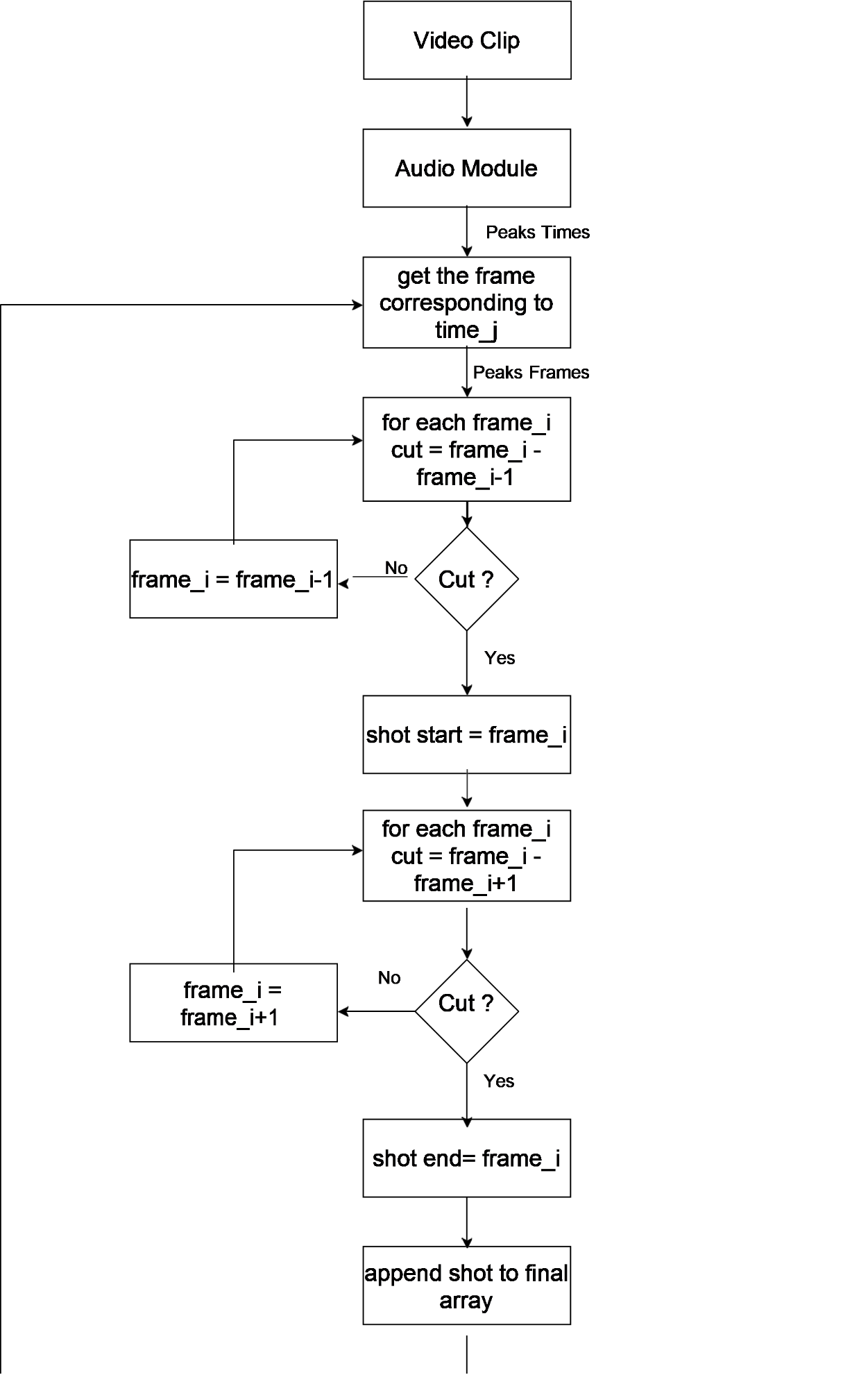
2- Starting from this frame explore frames ascendingly tell finding a shot cut

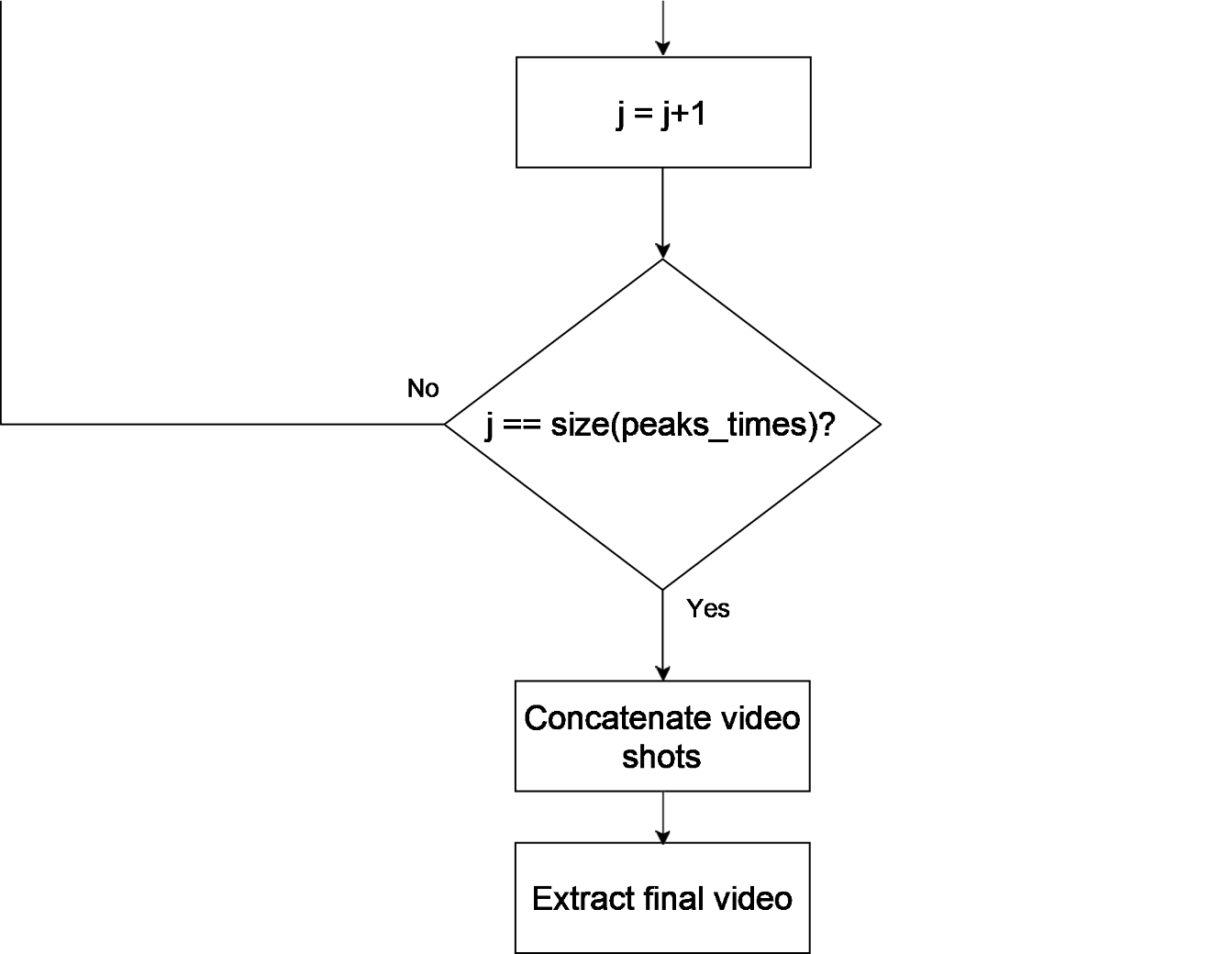
= shot end

3- Having shot start and shot end, append this shot to the final array

5- Concatenate shots into one array

6- Extract the summarized video

****

****

**Chapter 6: System Testing, Verification and Results**

Testing a complete system that has a lot of separate modules each has a functionality and some modeuls depend on the output of others is not only a difficult task but also illogical.So, Testing each phase invidually was the logical step to take and of the atmost importance the shot boundary phase because all other phases depend on the results of the shot boundary phase and also to ensure that each phase is working properly on it’s own before integrating all phases and testing the completed system.

In this space, before the first section, write an introductory paragraph on how you test and verify the correct operation of your system

**6.1. Testing Setup**

Explain the setup you are using in testing your project

To test such a system where the input is a video of a soccer match and the output is also a video but it’s the summary of said match, you have to have somekind of labels to verify the output wether that label is for each phase or a comlete summary to test the whole system. Unfortunately the labels are not available in our case so we had to resort to some manual verication methods and visual observation in almost all modules to ensure that they are working properly. Also, we developed our own score function to determine wether the output of the system is satisfactory or needs improvement.

**6.2. Testing Plan and Strategy**

Explain the methodology you follow while testing your project in details

we calculate a score function based on the goals detected, the interesting events detected and also the length of the output summarized video, goals is the most interesting event so we multiply the ratio of the detected goals and total number of goals in the match by 0.6, for other interesting events we multiply it’s ratio by 0.3 and we take the difference of 1 and ratio of length of output video and total length of video and multiply it by 0.1 to give a higher score for smaller length output video, then we add results to obtain the score.

**Score Function:**

**Score** = 0.6\***E1** + 0.3\***E2** + 0.1\*(1 - **E3**)

E1: no. Goals detected / total goals

E2: no. Of other interesting events detected / total no. Of events

E3: Length of output video / total length of video (=90)

**6.2.1. Module Testing**

Explain the steps you carried out to test different modules within the project. Give and discuss the results obtained from the testing of these modules

**Shot boundary testing and results**

The shot boundary algorithim was the first module developed and As mensioned before itis the most important because all other phases except the summarization depend on it’s output therefore it required a lot of testing and verification.

Unfortunately, there is not a source where a person can acquire a soccer match and the start and end of each shot in the match, that kind of information is not available. At least for free so we had to resort to visual observstion to verify the shot boundary algorithm output.

Extracting the video frames into disk and observing the cuts then comparing the output of the algorithm with what we see in the frames was the simplest and fastest way to improve the algorithm and refine the values of the thresholds used in it. The results were far better from what is anticipated

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Duration in minutes | Number of cuts | Efficiency |
| Test1.mp4 | 5 | 15 | 100% |
| Test2.mp4 | 2:22 | 21 | 80.9% |
| Test3.mp4 | 1:36 | 19 | 89.4% |
| Test4.mp4 | 3:12 | 38 | 84.2% |
| Test5.mp4 | 3:18 | 38 | 92% |
| Test6.mp4 | 4:41 | 37 | 86% |

**Audio version testing and results**

To find out which audio level is best for each league we had to try different matches for different leauges around the world with the three levels to determine which is the best suited for its respective league to produce the best summary.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| League | Match | Audio Level Threshold | Output video Length in minutes | Goals Detected accuracy | Other events detection accuracy | Score |
| BUNDESLIGA | Fc Augsburg  VS  Dortmund | 85 | 28:06 | 4/8 | 7/13 | 0.53 |
| 90 | 10:03 | 3/8 | 3/13 | 0.38 |
| 95 | 8:51 | 2/8 | 0/13 | 0.24 |
| Bayern Munich  VS   1. Bremen | 85 | 16:06 | 6/7 | 16/16 | 0.9 |
| 90 | 12:07 | 5/7 | 14/16 | 0.78 |
| 95 | 8:28 | 4/7 | 11/16 | 0.64 |
| Bayer Leverkusen  VS   1. Dusseldorf | 85 | 20:35 | 3/3 | 16/16 | 0.98 |
| 90 | 15:50 | 3/3 | 10/16 | 0.87 |
| 95 | 6:47 | 2/3 | 6/16 | 0.6 |
| LIGUE 1 | PSG  VS  Monaco | 85 | 31:14 | 6/6 | 9/12 | 0.89 |
| 90 | 23:46 | 6/6 | 10/12 | 0.92 |
| 95 | 14:48 | 5/6 | 7/12 | 0.76 |
| Rennes  VS  Nantes | 85 | 17:28 | 5/5 | 11/14 | 0.92 |
| 90 | 12:38 | 4/5 | 8/14 | 0.74 |
| 95 | 5:01 | 4/5 | 4/14 | 0.66 |
| PSG  VS  Marseille | 85 | 26:27 | 4/4 | 8/10 | 0.91 |
| 90 | 19:44 | 4/4 | 7/10 | 0.89 |
| 95 | 8:35 | 4/4 | 6/10 | 0.87 |
| PREMIER LEAGUE | Man. United  VS  Man. City | 85 | 29:38 | 2/2 | 5/8 | 0.85 |
| 90 | 21:42 | 2/2 | 4/8 | 0.83 |
| 95 | 10:33 | 2/2 | 2/8 | 0.76 |
| Aston Villa  VS  Man. City | 85 | 25:16 | 7/7 | 8/8 | 0.97 |
| 90 | 17:01 | 7/7 | 7/8 | 0.94 |
| 95 | 13:07 | 7/7 | 4/8 | 0.83 |
| Chelsea  VS  Arsenal | 85 | 24:29 | 4/4 | 7/10 | 0.88 |
| 90 | 18:04 | 4/4 | 6/10 | 0.86 |
| 95 | 6:51 | 3/4 | 3/10 | 0.63 |

In the above chart, for each league an average score is calculated for each audio threshold level by adding their scores and divide it by total number of matches (=3).

For most leagues we have tested level 85 threshold has the best results.

In premier league level 90 threshold has an average score very close to level 85 threshold so we can use it instead to reduce the length of the output summarized video.

**6.2.2. Integration Testing**

Explain the steps you carried out to test the integrated system of your project. Give and discuss the results obtained from this whole project testing

**6.3. Testing Schedule**

Mention your testing schedule

**6.4. Comparative Results to Previous Work**

Give a summary of comparative results to previous work in Tabulated and or Graphical form along with a short commentary.

**Chapter 7: Conclusions and Future Work**

We presented the main purpose of our system including our motivations, scope of work, problem definition and our proposed solutions. Then we provided the reader with necessary background information that is needed for subsequent chapter. After that we introduced the proposed system architecture in detail. In which we developed a new algorithm to detect the shot boundary. Afterwards, we discussed the most likely applied techniques in shot classification. i.e. categorize the detected video shots which came from shot boundary detection into five classes: wide, Medium, Close , Close-out , logo. logo detection is important in detecting important events. Replay segments are good indicators for exciting events in any sport video. We introduced a comprehensive foundation for event detection technique. Then solid concepts and techniques introduced a powerful set of tools for extracting features of interest such as audio processing and goal detection . After that we highlighted the concept of the summarization process as a collection of features which came from the preceding phases The reader of this project will reach a level of understanding which is the foundation of most of the work that is currently underway in the field of soccer summarizatiions.

**6.1. Faced Challenges**

* Dominant Color Detection

There are numerous variations which affect the degree of dominant color detection. We can mention the effect of changing the type of sport which led to change some thresholds according to the selected sport to truly detection. Also the stadium condition such as (lighting and shadows) can be considered as another reason which makes some troubles in the color region detection and we must be aware of the environmental factors such as (sunset, shadows and snow) and the field properties within the stadium which have worse effects than the mentioned factors [1].

* Shot boundary Detection the most significant factors that affect the accuracy of shot boundary detection because it is a critical stage in the whole system which all other stages follow it. Shot transition is an example of these factors especially when it corrupted by the occurrence of camera or object motion which make confusion in this process. The object motion can easily be fast because we are talking about a soccer game which means that all team players must move rapidly during the match that makes some troubles and cause wrong detection of shot boundary. In addition the movements of camera zooms present another factor that effect this process, and this is an essential operation that must be done to show different views of the game.
* Shot Classification
* Goal Mouth
* Goal detection
* Summarization

**6.2. Gained Experience**

Working on a project like this guarentees valuable experience and precious knowledge in the fields of computer vision, machine learning and image processing. Reasearching to find a solution for a hard problem such as shot boundary, goal mouth and shot classification is experience in of it self because you can apply the research process to any problem irrelevant of the field.

**6.3. Conclusions**

This area of soccer video processing is a dynamic field and new techniques and applications are reported routinely in professional literature and in new product announcements. For this reason, the topics included in this chapter served the purpose of introducing the state of art in soccer video summarization techniques. In addition to enhancement, this project served the purpose of improving a number of concepts such as shot boundary detection, shot classification and goal detection that can be used for other applications.

Laخsly is a fast and accurate soccer summarization system that is comprised of five phases, pre-processing phase in which we divide the video in a number of patches each patch is roughly 2000 frames to avoid memory overhead.

shot boundary detection phase segments the whole input video stream into small video shots by applying a combination of image processing techniques and histogram comparison techniques.

In the shot classification phase , the system applies different algorithms, namely; Grass dominant ratio extraction, Face Detection ,Deep learning and image processing techniques to classify the shot into one of the following classes :wide – medium- close – close out – logo. The logo classification is used for replay detection which is an important factor in determining which of the shots are important and which are not.

In The event detection phase each is shot in sconsedered seperatly and compared with the previous shot to determine if a goal has happened between the two shots by comparing the scoreboard in the upper left corner. Two methods are proposed to detect a goal event, Optical character recognition (OCR) and structural similarity image index (SSIM).audio loudness is detected as well as the goal mouth

Lastly, the summarization phase which takes into account the output of the previous phases to decide if he shot is important and shoul be included in the final video or should be discarded. For example, if a shot contains a goal then it is important or if a shot has a goal mouth and the audio level is high then it is important or it’s a replay shot and a replay of course is for something important. Then combining the important shots only in a video and that is the output of the system

There is even a faster version of laخsly that relies on ahot boundary and audio loudness only to produce a summary of the match in record time

There are a few limitations that are very hard to overcome in any sports video summarizations such as the cheer volume and variety in leagues and tournaments around the world which makes it very hard to create a global detection method for the score for example or a find a uniform way to detect important event therefore you have to tailor your system to a finite ste of league that have common features .

**6.4. Future Work**

this project discusses new implementations and techniques as well as existing ones in the area of soccer summarization.

Several of the areas that may be promising for further research is described in the following:

* Porting laخsly as a mobile application that can scale to large number of users, the existing web service should ease deployment and maintenance of the mobile version.
* Taking the advantage of using distributed system to improve the process time of our application, and use poor capabilities computers to collect their capabilities to distribute the work among them, without the need to buy a super computer to do the work if there are existing computers with low capabilities.
* Adapt the logo detection to work with more championships and different environments.
* Gathering more statistics about the played soccer game, and move from the soccer summarization area zone to the soccer analysis area zone with the use of new techniques such as object tracking.
* The research on finding relationships between audio and video features for soccer video analysis is a promising avenue. More research is based on the loudness, silence and pitch generated by a commentator and/or crowd are effective mea¬surements for detecting excitement.
* Extension of the proposed approach to different sports, such as American foot¬ball, basketball, and baseball, which require different event and object detection modules, will be addressed in the future.

**References**

The references should be ordered according to their appearance in the text. Ensure that all references are cited throughout your report text. The following are examples of how to write different types of references “[1] Book, [2] Journal/magazine articles, [3] conference paper, [4] website, [5] thesis”. Replace the fields with those of your used references. Question marks “??” should be replaced by the corresponding number

1. Author1, Author 2,…, “Book title,” name of publishing firm, edition, year
2. Author1, Author2,…., “Title of journal article,” name of the journal, vol. ??, no. ??, pp. ??, year of publication
3. Author1, Author2,…, “Title of conference paper,” in proceedings of conference name, city, country, date, year, pp. ??
4. Author or Corporation name, “Title,” year, link for the website, last accessed: date of last access
5. Author, “Thesis title,” M.Sc./Ph.D. thesis, Department, University, year

**Appendix A: Development Platforms and Tools**

This appendix explains used tools, platforms, and hardware kits. Any ready-made module should be mentioned and discussed in this appendix. The appendix is divided into two main sections; one for the hardware and the other is for software. Within each section, you could add as much subsections as needed, according to the number of tools and platforms that you use in your project.

In this space, before the first section, write an introductory paragraph to the appendix

**A.1. Hardware Platforms**

A description of any used hardware platforms/kit should be written in this section. Each platform/kit is better described in a separate subsection. (A1.1..)

**A.2. Software Tools**

Python Python is an interpreted, high-level, general-purpose programming language

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code.

Movie py MoviePy is a Python module for video editing, which can be used for basic operations (like cuts, concatenations, title insertions), video compositing (a.k.a. non-linear editing), video processing, or to create advanced effects. It can read and write the most common video formats

Keras – tensorflow

**Appendix B: Use Cases**

Include all your use cases

**Appendix C: User Guide**

Prepare a user guide for your project. Ensure that the guide is clear, detailed and easy for an ordinary customer to use your project. Employ figures and charts as needed to facilitate the use of your guide

**Appendix D: Code Documentation**

Your code or parts of the code you feel necessary could be included here (optional) however for one copy of this report an attached CD with all of the code is a must.

Remember you will deliver three copies of this report.

**Appendix D: Feasibility Study**

Give a detailed feasibility study of your project